Metabarcoding and Metagenomics 9: 621–635 (2025)
DOI: 10.3897/mbmg.9.156215

Short Communication

Constraining lake ecology via metagenomics, metatranscriptomics, and amplicon sequencing

Julia K. Nuy[1,2,3], Till L. V. Bornemann[1,2], Daniela Beisser[2,4], Alexander J. Probst[1,2], Jens Boenigk[2,3]

1 *Group of Environmental Metagenomics, University of Duisburg-Essen, Universitaetsstr. 5, Essen, Germany*
2 *Center of water and environmental research (ZWU), University of Duisburg-Essen, Universitaetsstr. 5, Essen, Germany*
3 *Department of Biodiversity, University of Duisburg-Essen, Universitaetsstr. 5, Essen, Germany*
4 *Department of Engineering and Natural Sciences, Westphalian University of Applied Science, Recklinghausen, Germany*

Corresponding author: Julia K. Nuy (Julia.nuy@uni-due.de)

Abstract

Lake ecosystems are hotspots for important ecosystem functions, yet linking their microbiome to physicochemical parameters remains a challenge. Here, we compare 16S rRNA gene-based amplicon sequencing, metagenomics, and metatranscriptomics across 21 European lakes using three strategies: (i) mapping shotgun reads to amplicon-derived OTUs, (ii) marker-specific profiling (rpS3 for metagenomes, rRNA reads for metatranscriptomes), and (iii) recovery of 16S rRNA genes from shotgun assemblies. Strategy (iii) proved unfeasible due to chimeric and highly variable assemblies and was excluded from further analyses. Both strategies (i) and (ii) revealed systematic methodological constraints. Amplicons yielded significantly lower richness and Shannon diversity than metagenomes and metatranscriptomes, while marker-based profiling highlighted broader detection of rare and active taxa. Despite these differences, all lakes showed the same relative ranking of diversity (metatranscriptomes > metagenomes > amplicons), indicating consistent methodological signatures across ecosystems. Beta-diversity analyses confirmed stronger concordance between metagenomes and metatranscriptomes than between either of these and amplicons. Differential abundance analyses further revealed method-specific detection biases, particularly for Proteobacteria and Bacteroidetes, that persisted even after correcting for 16S rRNA gene copy number and primer bias. Linking communities to physicochemical parameters, Mantel and Procrustes analyses showed the strongest global associations for metatranscriptomes, while metagenomes yielded the most stable explanatory OTUs in dbRDA and BioEnv models. Our results demonstrate that the sequencing approach is not merely a technical choice but represents an analytical dimension that fundamentally influences how microbiome–environment interactions are detected and interpreted.

Key words: Environment, microbiome, multiomics, OTUs, shotgun-sequencing, taxonomic profiling

OPEN ACCESS

Academic editor: Jianghua Yang
Received: 22 April 2025
Accepted: 29 October 2025
Published: 1 December 2025

Citation: Nuy JK, Bornemann TLV, Beisser D, Probst AJ, Boenigk J (2025) Constraining lake ecology via metagenomics, metatranscriptomics, and amplicon sequencing. Metabarcoding and Metagenomics 9: e156215. https://doi.org/10.3897/mbmg.9.156215

Copyright: © Julia K. Nuy et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Understanding the interaction between organismal data, such as lake microbiomes, and physicochemical lake ecosystem conditions is crucial for enhancing our comprehension of ecological processes. While it is undeniable

that the composition of a microbiome is tightly linked with physicochemical ecosystem conditions, it remains uninvestigated whether commonly applied methods perform comparably in reflecting environmental features (Suppl. material 3: table S1). Currently, most microbiome studies rely on next-generation sequencing data, generated from community DNA or RNA. Widely used and highly time and cost-effective amplicon sequencing techniques primarily screen microbiomes by targeting specific hypervariable regions of marker genes, such as the 16S rRNA gene for prokaryotes. PCR-amplification enables the detection of the rare biosphere (Sogin et al. 2006), whereas amplification-free metagenomes provide genomic blueprints of the dominantly occurring community members. Metatranscriptomes, being the most degradation-sensitive material, provide a snapshot of the metabolically active community (protein synthesizing (Blazewicz et al. 2013), not necessarily growing) by capturing ribosomal and messenger RNA (Mills et al. 2012) (rRNA and mRNA, respectively).

The massive generation of metagenome-assembled genomes has significantly expanded databases, enabling taxonomic profiling using metagenomes and assembled metagenomes, and allowing for comparisons of amplicon sequences and metagenomic shotgun technologies. However, taxonomic profiles derived from different markers and databases can vary due to methodological and database-specific biases. Even when focusing on a single marker (e.g. 16S rRNA gene), systematic biases can arise depending on the method of sequence recovery (e.g., amplicon sequencing, and assembly of metagenomes and metatranscriptomes) as these approaches differ in sensitivity, coverage and susceptibility to assembly or amplification artifacts. While such methodological biases are well recognized, it has not been systematically evaluated how they influence ecological interpretability, specifically the robustness of associations between microbiome composition and environmental features in lake ecosystems.

To address this issue, we tested three strategies: First, we mapped shotgun reads to amplicon-derived OTUs (strategy I; OTU-based). Second, we applied marker-specific profiling, using the most informative taxonomic marker for each dataset e.g., single copy rpS3 for metagenomes and rRNA reads for metatranscriptomes (ii strategy; marker-based). Third, we attempted to recover 16S rRNA genes from shotgun assemblies, where applicable, to build a cross-method OTU library (strategy iii).

However, strategy iii proved unsuccessful, as assemblies yielded highly heterogeneous lengths and chimeric sequences exceeding 1600 bp (Suppl. material 1: fig. S1), consistent with previous reports (Nelson and Stegen 2015; Yuan et al. 2015). This prevented a meaningful comparison with the V9 region of the amplicon OTUs, and we therefore did not pursue this approach further.

In the following, we focus on the resulting community profiles of strategy i and ii to assess consistency and methodological biases. Unlike most studies, which typically model community composition as a function of environmental variables (Community ~ Environment), we specifically focus on the reverse perspective (Environment ~ Community) to improve the comparability of different sequencing approaches and taxonomic markers. To our knowledge, this is the first comprehensive comparison of amplicon, metagenomic, and metatranscriptomic sequencing in lake microbiomes with regard to their potential to reflect and predict environmental features.

For this purpose, we generated a dataset for 21 lakes located across Europe encompassing 16S rRNA gene-based amplicon sequencing data, as well as metagenomic and metatranscriptomic sequences (Shah et al. 2024), encompassing 3.31 Gbps, 170.91 Gbps, and 34.11 Gbps of sequencing depth, respectively (Suppl. material 1: figs S2, S3). To address the marker and database biases mentioned above we pursued the two different strategies. First (strategy (i)), we mapped the metagenomic and metatranscriptomic reads to a database consisting of representative OTUs from the amplicon dataset that passed a relative read abundance threshold of 0.001% per sample (Nuy et al. 2020). This filtering step resulted in 2535 OTUs recruiting amplicon 9,356,192 reads, which collectively recruited 154,996 metagenomic reads and 19,979,498 metatranscriptomic reads (see Suppl. material 2 for information about metatranscriptome, metagenome, and amplicon data) (Sharon et al. 2015). Each OTU recruited metagenomic and metatranscriptomic reads across all samples; consequently, the technologies used, represented the same 26 prokaryotic phyla. Despite this consistency in phylum coverage, alpha diversity metrics differed substantially between methods (Fig. 1A). Friedman tests confirmed highly significant differences ($p < 0.001$; Kendall's $W = 0.76$ for richness and 0.84 for Shannon), with pairwise Wilcoxon tests showing that amplicon data yielded significantly lower richness and Shannon values compared to both metagenomes and metatranscriptomes, while no significant differences were observed between metagenomes and metatranscriptomes (Suppl. material 3: table S2).

In strategy (ii), we applied marker-specific taxonomic profiling, i.e. rpS3 single-copy genes for metagenomes and mapped rRNA reads for metatranscriptomes, while retaining amplicon OTUs as an independent reference. This approach led to 23 prokaryotic phyla being represented in metagenomes and 51 phyla in metatranscriptomes (Fig. 1B). Here, alpha diversity differences were even more pronounced, with metatranscriptomes yielding the highest richness and Shannon diversity, followed by metagenomes and amplicons. Friedman tests were again highly significant ($p < 0.001$), with Kendall's $W = 0.862$ for richness and 0.857 for Shannon. Although absolute diversity values differed markedly between sequencing approaches, all lakes exhibited a consistent relative ordering of methods (metatranscriptomes > metagenomes > amplicons) (Suppl. material 3: table S2). This stability across ecosystems indicates that the observed differences are not lake-specific, but rather reflect systematic methodological effects.

Previous studies have reported conflicting results regarding the detection of phyla with different sequencing approaches, which is also a consequence of comparing different taxonomic markers (Poretsky et al. 2014; Guo et al. 2016; Tessler et al. 2017; Brumfield et al. 2020; Khachatryan et al. 2020; Hempel et al. 2023), as also presented in Fig. 1. Some studies indicated that less than 50% of phyla are detected with metagenomes compared to amplicons (Tessler et al. 2017) and reveal a poorer taxonomic resolution, while others demonstrated the exact opposite (Poretsky et al. 2014; Guo et al. 2016; Tessler et al. 2017; Brumfield et al. 2020; Khachatryan et al. 2020; Hempel et al. 2023). However, comparisons of short reads to public databases remain challenging; for instance, 150 bp reads only cover approximately 15% of the average prokaryotic gene (Xu et al. 2006), leading to inherent biases in taxonomic calling.

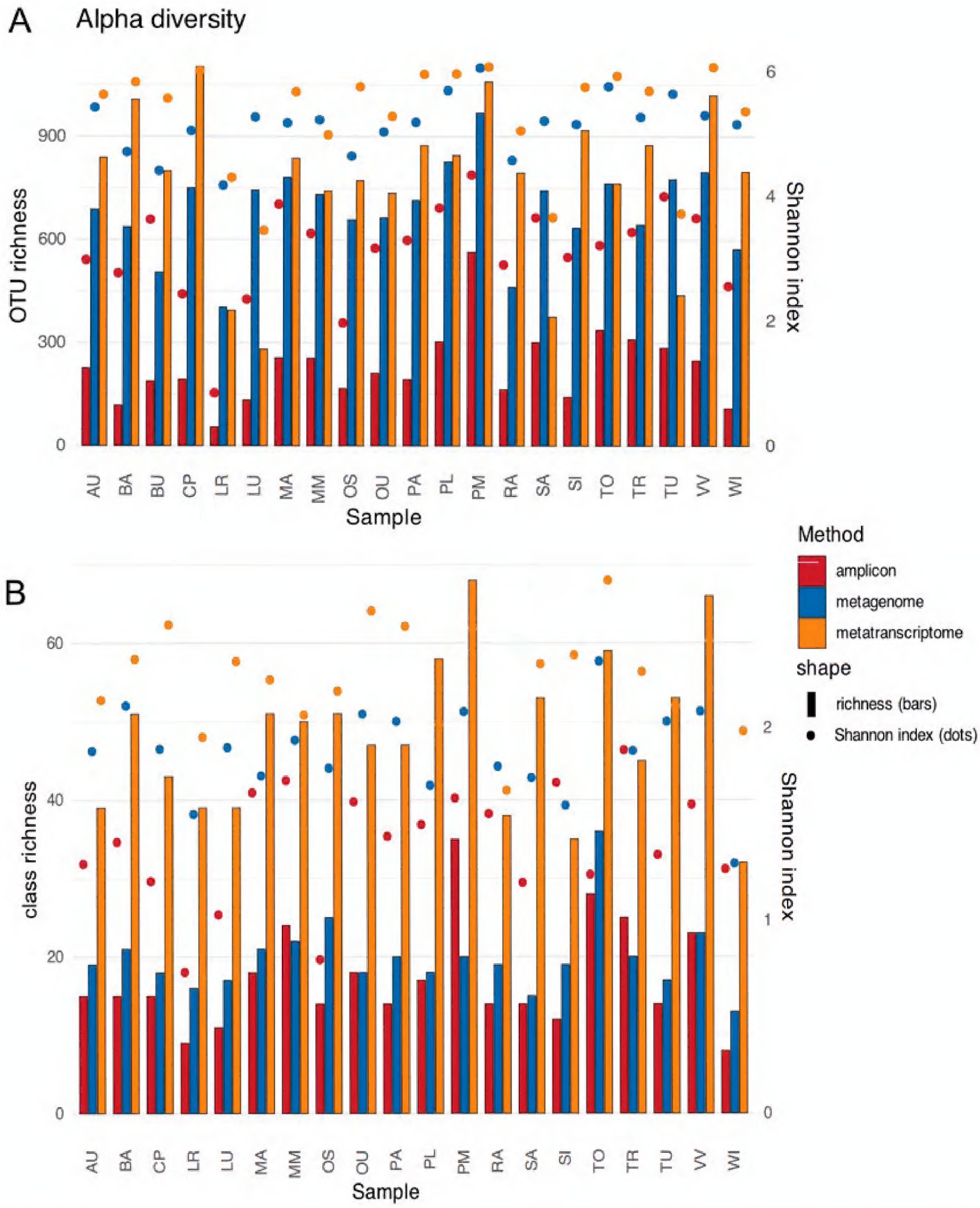

Figure 1. Alpha diversity comparison across sequencing approaches. (**A**) OTU richness (bars) and Shannon diversity index (dots) per sample and (**B**) Class richness (bars) and Shannon diversity index (dots) per sample. Colors represent sequencing approaches: amplicon (red), metagenome (blue), and metatranscriptome (orange). Diversity estimates show systematic differences across methods, with metatranscriptomes generally yielding highest richness estimates.

To minimize such database-related biases, we pursued two complementary approaches. At the OTU level, mapping metagenomic and metatranscriptomic reads to amplicon-derived OTUs allowed a consistent comparison across methods. At the marker level, we retained the most informative marker for each technology (rpS3 for metagenomes, rRNA reads for metatranscriptomes, amplicon OTUs as reference), reducing taxonomic resolution to the class level for comparability.

Both strategies consistently revealed systematic methodological effects. PERMANOVA indicated that the sequencing approach explained ~16% of the variance in relative abundance data and ~13% for presence-absence data. Beta-dispersion analyses further showed that metatranscriptomes produced the most homogeneous profiles, whereas amplicons were the most heterogeneous (Suppl. material 3: table S2). At the marker level, method effects were even stronger (~41% explained variation). Presence-absence and abundance-based OTU models differed by only ~3%, suggesting that the main bias arises from taxon detection rather than abundance estimates. For this reason, we focus primarily on abundance-based analyses, while presence-absence results are provided in the Suppl. materials.

Procrustes and Mantel tests further confirmed stronger concordance between metatranscriptomes and metagenomes than between either of these and amplicons. Correlations were consistently higher for relative abundance than for presence-absence data (Fig. 2, Suppl. material 1: figs S4–S6; Suppl. material 3: tables S2, C, S3).

Together, the results demonstrate that method choice systematically shapes community profiles. Differences between presence-absence and abundance-based data were minor, whereas marker-level comparisons highlighted stronger discrepancies in detection breadth and abundance structure.

The main differences between metagenomes and amplicon sequencing data were attributable to different abundances of taxa that recruited low to average numbers of reads, as reported previously (Poretsky et al. 2014). Differential abundance analysis of OTUs revealed a higher detection of OTUs affiliated to Alphaproteobacteria, and Cyanobacteria in amplicons compared to metagenomes, whereas Gammaproteobacteria (mostly Burkholderiales) were underrepresented (Fig. 3, Suppl. material 3: table S4). Specifically, *Polynucleobacter*, and *Limnohabitans* were strongly underrepresented in the amplicon dataset, a pattern also observed in a study of the same lakes analyzed with the same amplicon data and attributed to primer biases (Nuy et al. 2020)(Suppl. material 1: fig. S7). At the same time, certain taxa associated with Bacteroidetes and Cyanobacteria were differentially represented across both methods. In Bacteroidetes, metagenomes capture higher relative abundances of Cytophagales (Algoriphagus) and Chitinophagales (Dingihuibacter), whereas amplicons showed an overrepresentation of Flavobacteriales (Flavobacterium). The in OTU-based strategy i observed patterns were partially consistent with the marker-based approach at the class level. At this aggregated level, patterns were largely consistent with OTU-based analyses, and Chi^2 tests of class distributions did not reveal significant deviations between methods (Suppl. material 1: fig. S8). This suggests that while fine-scale differences in detection are masked, taxonomic aggregation enhances comparability across heterogeneous markers.

While the relative abundances of assembled metagenomes have been shown to correlate significantly with quantitative digital droplet PCR measurements (Probst et al. 2018), amplicon data suffer from primer bias (Probst et al. 2015; Starke and Morais 2019), amplification biases (Acinas et al. 2005; Stach et al. 2023), chimeras (Haas et al. 2011), and variable numbers of 16S rRNA gene copies per genome (Farrelly et al. 1995). To test if these biases significantly affect the amplicon data *in silico*, we corrected

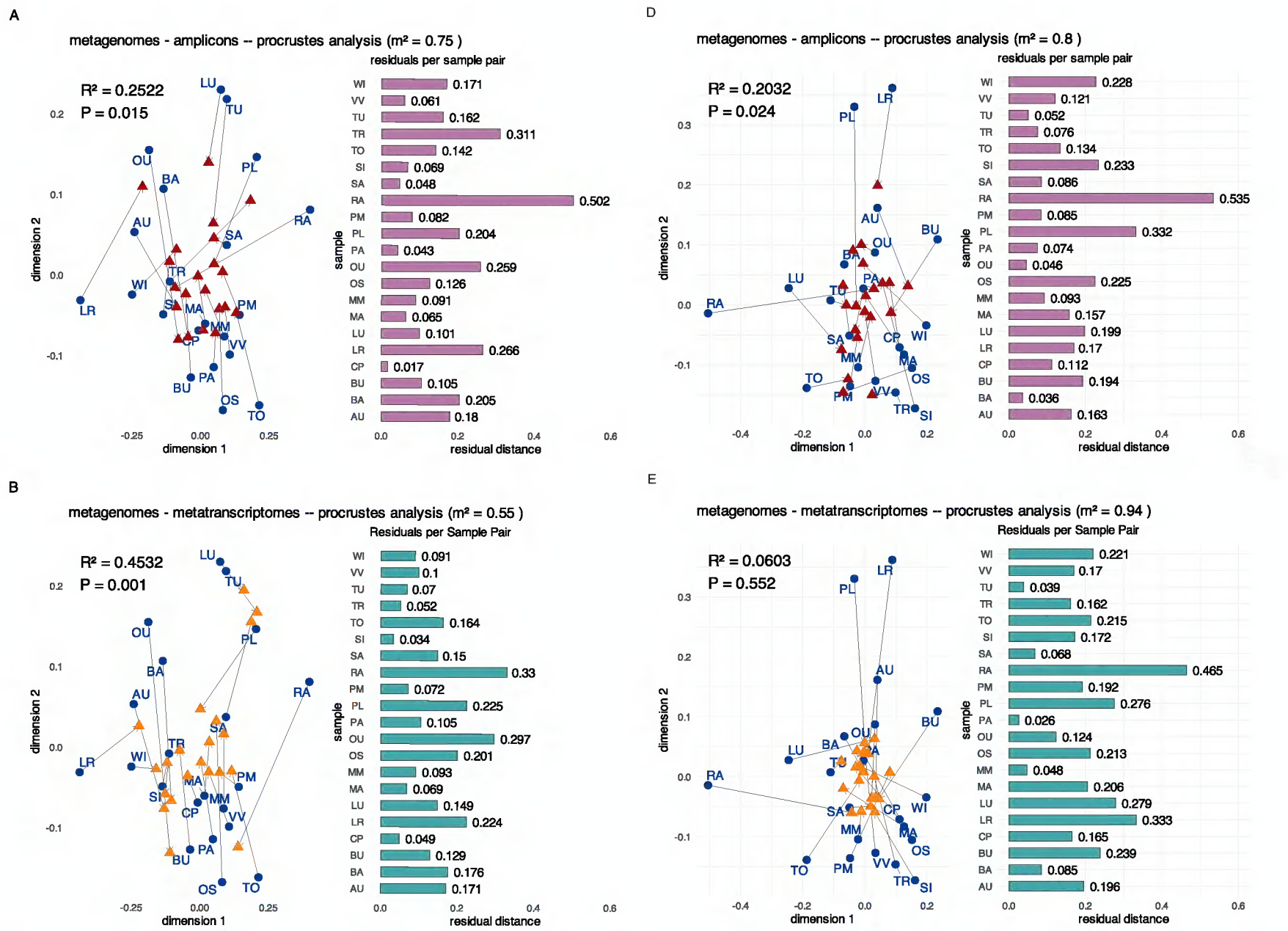

Figure 2. Procrustes analyses comparing metagenomes, metatranscriptomes, and amplicon data. The Procrustes analyses were performed to assess congruence between metagenomes, metatranscriptomes, and amplicons. Reference datasets are shown as circles, target datasets as triangles, and lines connect paired samples, with line length indicating the degree of dissimilarity. The accompanying barplots quantify residual distances per sample pair, with colors indicating the respective comparison (purple = metagenomes vs. amplicons, dark orange = amplicons vs. metatranscriptomes, teal = metagenomes vs. metatranscriptomes). OTU-based comparisons (**A–C**): **A.** Metagenomes (reference; blue) vs. amplicons (target; red). **B.** Metagenomes (reference; blue) vs. metatranscriptomes (target; yellow). **C.** Amplicons (reference; red) vs. metatranscriptomes (target; yellow). Marker-based comparisons (**D–F**; 16S rRNA genes and rpS3): **D.** Metagenomes vs. amplicons. **E.** Metagenomes vs. metatranscriptomes. **F.** Amplicons vs. metatranscriptomes. The Procrustes correlation coefficient (m^2) quantifies similarity between ordinations, with values closer to 0 indicating higher congruence. The R^2 value represents the proportion of variation in one dataset explained by the other. Among OTU-based comparisons, the strongest congruence was observed between metagenomes and metatranscriptomes (R^2 = 0.68, P = 0.001), whereas amplicons showed greater divergence from both other methods. Marker-based comparisons yielded overall higher congruence, with stronger alignment between methods and lower residual distances across samples, indicating that marker-based approaches capture shared community structure more consistently than OTU-based analyses.

the dataset for 16S rRNA gene copy numbers (Stoddard et al. 2015) and the specific primer bias by using a weighted primer score (Starke and Morais 2019) (Suppl. material 1: fig. S7). Correcting relative abundances by both gene copy number and weighted primer score did not significantly alter community composition (uncorrected amplicon dataset vs. 16S copy number correction: r_s = 0.9925, p = 0.001; uncorrected amplicon dataset vs weighted primer score correction: r_s = 0.9968, p = 0.001), indicating that amplification biases (Suzuki and Giovannoni 1996; Polz and Cavanaugh 1998; Stach et al. 2023) (in combination with a primer bias (Suppl. material 1: fig. S7)) are the major components distorting community profiles derived from lake amplicon data.

The active community reflected by metatranscriptomes, as determined by the Procrustes analysis and Mantel test, was significantly similar to communities assessed with the metagenomic approach (Fig. 2, Suppl. material 1: figs S5, S6). Only six OTUs were detected to be differentially abundant (Fig. 3). Relating the relative abundance of taxa detected in metatranscriptomes to DNA-based methods (amplicon data or metagenome data) not only underscores methodological biases but also provides ecologically meaningful insights into microbial activity at the time of sampling. When comparing metatranscriptomes with DNA-based references, the choice of baseline method mattered. Among all classes, 22 classes exhibited similar proportions between metatranscriptomes and metagenomes, while the proportions of other classes differed (Suppl. material 1: fig. S9). For example, metatranscriptomes linked to metagenomes indicated that Clostridia, particularly *Clostridium* spp., of which some representatives are routine key indicators of fecal contamination and water quality, were predominantly active (Riđanović and Riđanović 2017; Stelma 2018; Li et al. 2025). In contrast, when using amplicon data as a baseline, a smaller fraction of this population transcribed 16S rRNA genes. These results suggest that the selection of the DNA profiling method as a reference to relate metatranscriptomes to, influences the inferences regarding the active members of a microbiome.

To comprehensively evaluate how sequencing strategies reflect environmental conditions, we applied four complementary community ecology approaches on the three datasets i.e. Mantel tests and Procrustes analyses to evaluate overall concordance between community and environmental distance matrices, BioEnv to identify the subset of environmental variables best matching community dissimilarities, and dbRDA with stepwise selection (ordistep, 999 permutations) to determine which variables significantly explain environmental variables (Fig. 4, Suppl. material 3: tables S3, S5, S6). In line with previous studies (Crump et al. 2007; Souffreau et al. 2015), we assume physicochemical factors and community composition are strongly linked. Therefore, we expect that microbiome communities explain a high proportion of variance in physicochemical factors.

In general, the Mantel tests and Procrustes analyses (Fig. 4) revealed a strong link between the physicochemical matrix and the metagenomic and metatranscriptomic dataset, while the amplicon dataset showed a low correlation in the Mantel test and a high measure of dissimilarity in the Procrustes analysis (Fig. 4, Suppl. material 3: table S5).

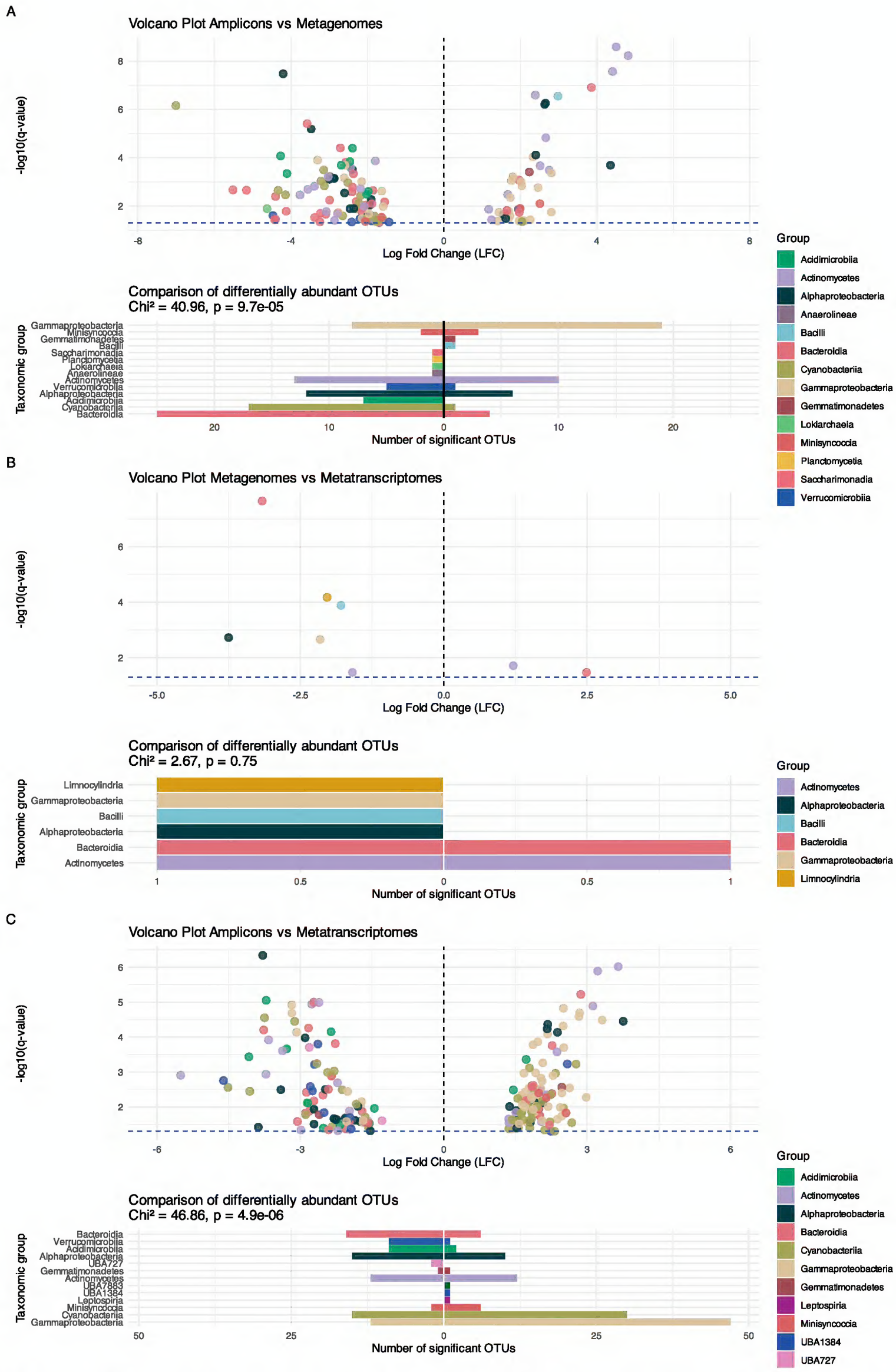

Figure 3. OTU-based differential abundance analysis across datasets. The volcano plots and corresponding bar charts illustrate the differential abundance of operational taxonomic units (OTUs) in **A**) amplicons vs metagenomes **B**) amplicons vs metatranscriptomes, and **C**) metagenomes vs. metatranscriptomes. Each comparison highlights taxonomic groups (refer to the color code in the legend) that are significantly over- or underrepresented in the respective datasets shown by the logfold change indicating the enrichment in one dataset over the other. The points and bars represent OTUs colored by their taxonomic classification. The accompanying bar plot quantifies the number of significantly different OTUs within major taxonomic groups. Enriched for each comparison can be found in Suppl. material 3: table S4. The marker-based analysis can be found in Suppl. material 1: fig. S8.

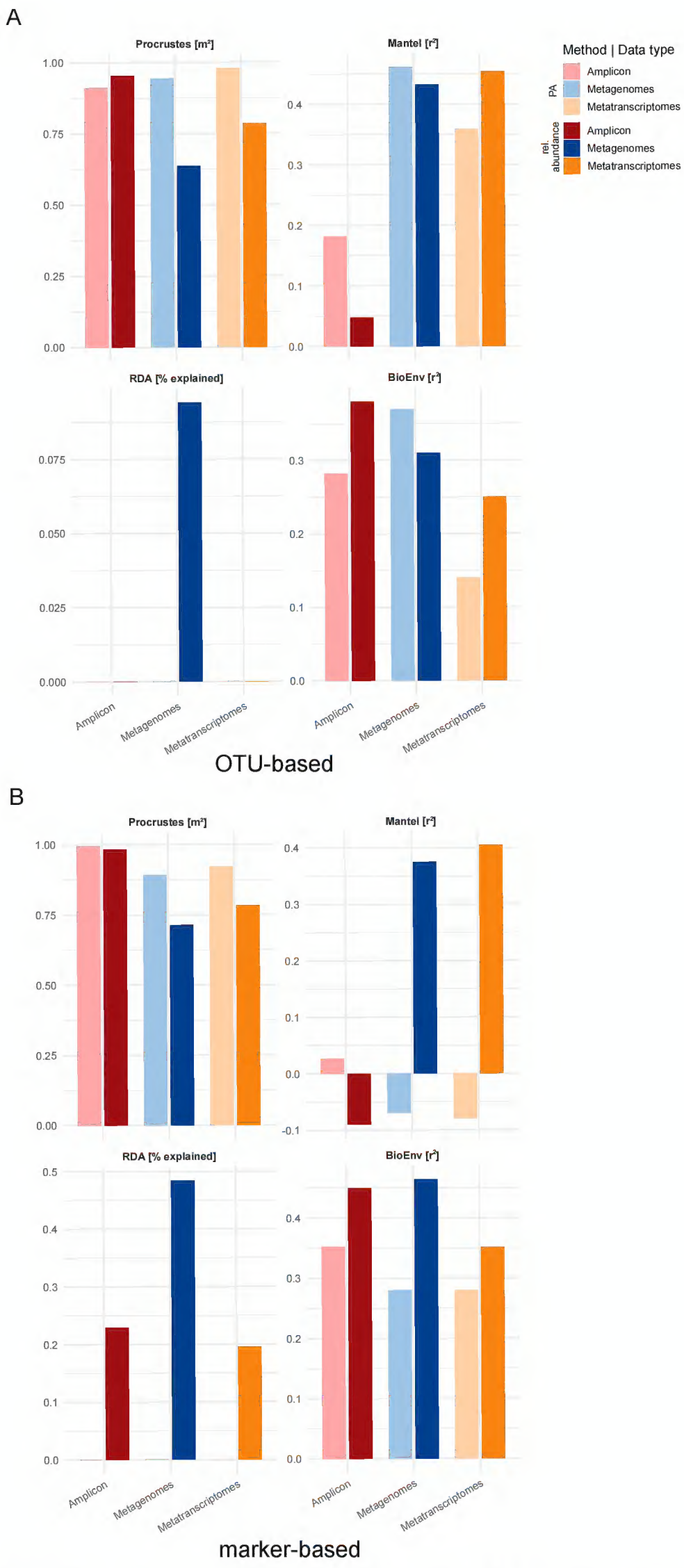

Figure 4. Comparison of ecological methods in reflecting environmental variation across OTU-based and marker-based datasets. Barplots summarize the performance of four different ecological analytical methods, including Procrustes analysis, Mantel test, Redundancy Analysis (RDA; explained variance), and BioEnv (Spearman's ρ; rs), across three datasets based on OTUs (**A**) and different taxonomic markers (**B**) represented by different colours. (amplicon = red; metagenome = blue; metatranscriptome = yellow). The y-axis represents the m^2 for the Procrustes analyses (ranging from 0 (highest similarity) and 1 (highest dissimilarity)), the spearman rank correlation coefficent r_s for the Mantel test and the BioEnv, and the fraction of explained variation in the RDA where higher values indicate a stronger relationship or explanatory capability. Across methods, marker-based data generally yielded stronger and more consistent relationships with environmental variables than OTU-based data. Underlying OTU and marker models used for RDA and BioEnv analyses are provided in Suppl. material 3: table S4.

In the dbRDA analyses, the initial full metagenomic model explained ~20% of environmental variation. However, stepwise forward selection reduced the model to a single significant OTU affiliated with Cyanobacteria, which alone explained ~9.4% of the variation (adjusted R^2 = 0.094, F = 3.08, p = 0.021). Neither the amplicon nor the metatranscriptomic datasets yielded significant models after stepwise selection. This holds also true for using presence-absence-based distances, suggesting that abundance information is essential to capture environmental gradients. BioEnv results mirrored this pattern, with metagenomes and amplicons showing the strongest match to environmental gradients. Importantly, the OTUs retained in BioEnv models differed taxonomically across presence-absence and relative abundance data and across methods, suggesting that each sequencing approach highlights different ecological aspects (Fig. 4, Suppl. material 3: tables S4, S5). At both OTU- and marker-based levels, the relative ranking of methods was stable: metagenomes and metatranscriptomes consistently outperformed amplicons. Yet, their strengths differed, metatranscriptomes captured global community-environment associations most strongly in Mantel and Procrustes tests, whereas metagenomes yielded the most stable explanatory OTUs in dbRDA and BioEnv. Amplicons, despite their low global concordance, still contributed individual OTUs that matched environmental gradients.

Our study across 21 European lakes demonstrates that sequencing strategy strongly shapes ecological interpretation. Shotgun data best captured global environmental patterns in Mantel and Procrustes analyses, while amplicon and metagenomic data provided more specific predictors in BioEnv and RDA, with metagenomes yielding the only stable OTU-level indicator. Each method entails distinct trade-offs: amplicon sequencing is rapid and cost-efficient but limited by primer bias and reduced resolution of environmental variability. Metatranscriptomics reveals the active community but is inherently variable due to transcriptional dynamics and sequencing depth, and its susceptibility to RNA degradation. Metagenomes provide the broadest genomic context and most stable explanatory OTUs, albeit at high sequencing and computational cost and with reduced sensitivity to detect the rare biosphere. By combining OTU-based normalization across all datasets with marker-specific profiling, we show that both strategies converge on the same conclusion: sequencing approach systematically structures community profiles and their ecological interpretation. OTU-based analyses emphasized comparability across datasets, while marker-specific analyses highlighted how detection breadth and abundance patterns differ by sequencing technology. Although controlled experiments such as batch cultures would be needed to ultimately determine the “best” approach, our results highlight that no single method is universally superior. Instead, each sequencing strategy comes with distinct strengths and limitations that directly shape ecological conclusions, underscoring the importance of methodological awareness when interpreting microbiome-environment relationships.

Our results demonstrate that the choice of sequencing approach extends beyond technical considerations and strongly influences ecological interpretation.

Acknowledgments

We thank the German Research Foundation (DFG) for the funding of the projects BO 3245/19 and BO3245/17. Further, we thank the team for publication services of the library of the University of Duisburg - Essen for providing funds for the open access publication fee.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

This work was supported by the German Research Foundation (DFG) for the funding of the projects BO 3245/19 and BO3245/17.

Author contributions

Conceptualization: JKN. Formal analysis: JKN. Software: TLVB. Supervision: JB, AJP, DB. Visualization: JKN. Writing - original draft: JKN. Writing - review and editing: TLVB, DB, AJP, JB.

Author ORCIDs

Julia K. Nuy https://orcid.org/0000-0003-0270-0070
Till L. V. Bornemann https://orcid.org/0000-0003-3182-6596
Daniela Beisser https://orcid.org/0000-0002-0679-6631
Alexander J. Probst https://orcid.org/0000-0002-9392-6544
Jens Boenigk https://orcid.org/0000-0001-8858-8889

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. Applied and Environmental Microbiology 71(12): 8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013) Evaluating rRNA as an Indicator of Microbial Activity in Environmental Communities: Limitations and Uses. The ISME Journal 7(11): 2061–2068. https://doi.org/10.1038/ismej.2013.102

Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB (2020) Microbial Resolution of Whole Genome Shotgun and 16S Amplicon Metagenomic Sequencing Using Publicly Available NEON Data. PLoS ONE 15(2): e0228899. https://doi.org/10.1371/journal.pone.0228899

Crump RC, Adams HE, Hobbie JE, Kling GW (2007) Biogeography of Bacterioplankton in Lakes and Streams of an Arctic Tundra Catchment. Ecology 88(6): 1365–1378. https://doi.org/10.1890/06-0387

Farrelly V, Rainey FA, Stackebrandt E (1995) Effect of Genome Size and Rrn Gene Copy Number on PCR Amplification of 16S rRNA Genes from a Mixture of Bacterial Species. Applied and Environmental Microbiology 61(7): 2798–2801. https://doi.org/10.1128/aem.61.7.2798-2801.1995

Guo J, Cole JR, Zhang Q, Brown CT, Tiedje JM (2016) Microbial Community Analysis with Ribosomal Gene Fragments from Shotgun Metagenomes. Applied and Environmental Microbiology 82(1): 157–166. https://doi.org/10.1128/AEM.02772-15

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW (2011) Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons. Genome Research 21(3): 494–504. https://doi.org/10.1101/gr.112730.110

Hempel CA, Buchner D, Mack L, Brasseur MV, Tulpan D, Leese F, Steinke D (2023) Predicting Environmental Stressor Levels with Machine Learning: A Comparison between Amplicon Sequencing, Metagenomics, and Total RNA Sequencing Based on Taxonomically Assigned Data. Frontiers in Microbiology 14(November): 1217750. https://doi.org/10.3389/fmicb.2023.1217750

Khachatryan L, de Leeuw RH, Kraakman MEM, Pappas N, te Raa M, Mei H, de Knijff P, Laros JFJ (2020) Taxonomic Classification and Abundance Estimation Using 16S and WGS—A Comparison Using Controlled Reference Samples. Forensic Science International. Genetics 46(May): 102257. https://doi.org/10.1016/j.fsigen.2020.102257

Li J, Sun W, Cao Y, Wu J, Duan L, Zhang M, Luo X, Deng Q, Peng Z, Mou X, Li W, Wang P (2025) Increased Temperature Enhances Microbial-Mediated Lignin Decomposition in River Sediment. Microbiome 13(1): 89. https://doi.org/10.1186/s40168-025-02076-z

Mills HJ, Reese BK, Shepard A, Riedinger N, Dowd SE, Morono Y, Inagaki F (2012) Characterization of Metabolically Active Bacterial Populations in Subseafloor Nankai Trough Sediments above, within, and below the Sulfate–Methane Transition Zone. Frontiers in Microbiology 3(April), 1–12. https://doi.org/10.3389/fmicb.2012.00113

Nelson WC, Stegen JC (2015) The Reduced Genomes of Parcubacteria (OD1) Contain Signatures of a Symbiotic Lifestyle. Frontiers in Microbiology 6: 713. https://doi.org/10.3389/fmicb.2015.00713

Nuy JK, Hoetzinger M, Hahn MW, Beisser D, Boenigk J (2020) Ecological Differentiation in Two Major Freshwater Bacterial Taxa Along Environmental Gradients. Frontiers in Microbiology 11(February): 154. https://doi.org/10.3389/fmicb.2020.00154

Polz MF, Cavanaugh CM (1998) Bias in Template-to-Product Ratios in Multitemplate PCR. Applied and Environmental Microbiology 64(10): 3724–3730. https://doi.org/10.1128/AEM.64.10.3724-3730.1998

Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT (2014) Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. PLoS ONE 9(4): e93827. https://doi.org/10.1371/journal.pone.0093827

Probst AJ, Weinmaier T, DeSantis TZ, Santo Domingo JW, Ashbolt N (2015) New Perspectives on Microbial Community Distortion after Whole-Genome Amplification. PLoS ONE 10(5): e0124158. https://doi.org/10.1371/journal.pone.0124158

Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF (2018) Differential Depth Distribution of Microbial Function and Putative Symbionts through Sediment-Hosted Aquifers in the Deep Terrestrial Subsurface. Nature Microbiology 3(3): 328–336. https://doi.org/10.1038/s41564-017-0098-y

Riđanović, L, Riđanović S (2017) The Impact of Climatic Factors on Bacterial Indicators of Freshwater Quality. Journal of Survey in Fisheries Sciences, 29–37. https://doi.org/10.18331/SFS2017.4.1.4

Shah M, Bornemann TLV, Nuy JK, Hahn MW, Probst AJ, Beisser D, Boenigk J (2024) Genome-Resolved Metagenomics Reveals the Effect of Nutrient Availability on Bacterial Genomic Properties across 44 European Freshwater Lakes. Environmental Microbiology 26(6): e16634. https://doi.org/10.1111/1462-2920.16634

Sharon I, Kertesz M, Hug LA, Pushkarev D, Blauwkamp TA, Castelle CJ, Amirebrahimi M, Thomas BC, Burstein D, Tringe SG, Williams KH, Banfield JF (2015) Accurate, Multi-Kb Reads Resolve Complex Populations and Detect Rare Microorganisms. Genome Research 25(4): 534–543. https://doi.org/10.1101/gr.183012.114

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial Diversity in the Deep Sea and the Underexplored 'Rare Biosphere.'. Proceedings of the National Academy of Sciences of the United States of America 103(32): 12115–12120. https://doi.org/10.1073/pnas.0605127103

Souffreau C, Van der Gucht K, van Gremberghe I, Kosten S, Lacerot G, Lobão LM, de Moraes Huszar VL, Roland F, Jeppesen E, Vyverman W, De Meester L (2015) Environmental Rather than Spatial Factors Structure Bacterioplankton Communities in Shallow Lakes along a > 6000 Km Latitudinal Gradient in South America. Environmental Microbiology 17(7): 2336–2351. https://doi.org/10.1111/1462-2920.12692

Stach TL, Sieber G, Shah M, Simon SA, Soares A, Bornemann TLV, Plewka J, Künkel J, Becker C, Meyer F, Boenigk J, Probst AJ (2023) Temporal Disturbance of a Model Stream Ecosystem by High Microbial Diversity from Treated Wastewater. MicrobiologyOpen 12(2): e1347. https://doi.org/10.1002/mbo3.1347

Starke R, Morais D 2019. "Gene Copy Normalization of the 16S rRNA Gene Cannot Outweigh the Methodological Biases of Sequencing." Preprint, bioRxiv, October 21. https://doi.org/10.1101/813477

Stelma Jr GN (2018) Use of Bacterial Spores in Monitoring Water Quality and Treatment. Journal of Water and Health 16(4): 491–500. https://doi.org/10.2166/wh.2018.013

Stoddard S, Smith B, Hein R, Roller RKB, Schmidt MT (2015) rrnDB: Improved Tools for Interpreting rRNA Gene Abundance in Bacteria and Archaea and a New Foundation for Future Development." Nucleic Acids Research 43(D1): D593–98. https://doi.org/10.1093/nar/gku1201

Suzuki MT, Giovannoni SJ (1996) Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. Applied and Environmental Microbiology 62(2): 625–630. https://doi.org/10.1128/aem.62.2.625-630.1996

Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, Segovia BT, Lansac-Toha FA, Lemke M, DeSalle R, Mason CE, Brugler MR (2017) Large-Scale Differences in Microbial Biodiversity Discovery between 16S Amplicon and Shotgun Sequencing. Scientific Reports 7(1): 6589. https://doi.org/10.1038/s41598-017-06665-3

Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW (2006) Average Gene Length Is Highly Conserved in Prokaryotes and Eukaryotes and Diverges Only Between the Two Kingdoms. Molecular Biology and Evolution 23(6): 1107–1108. https://doi.org/10.1093/molbev/msk019

Yuan C, Lei J, Cole J, Sun Y (2015) Reconstructing 16S rRNA Genes in Metagenomic Data. Bioinformatics (Oxford, England) 31(12): i35–i43. https://doi.org/10.1093/bioinformatics/btv231

Supplementary material 1

Additional figures

Authors: Julia K. Nuy

Data type: docx

Explanation note: **fig. S1.** Length distributions of recovered 16S rRNA gene centroids from metagenomes and metatranscriptomes. **fig. S2.** Rarefaction curves showing sequencing depth and species richness for OTU-based analyses. **fig. S3.** Rarefaction curves for marker-based profiling of metatranscriptomes and metagenomes. **fig. S4.** Procrustes analyses comparing microbial community profiles across sequencing methods. **fig. S5.** Relative abundance heatmaps of prokaryotic classes across European lakes for different sequencing strategies. **fig. S6.** Leave-one-out sensitivity analyses of Procrustes correlations across sequencing approaches. **fig. S7.** Primer bias analysis and 16S rRNA gene copy numbers across major bacterial phyla and classes. **fig. S8.** Differential abundance of microbial classes across sequencing approaches. **fig. S9.** Class-wise correlations of relative abundances across sequencing methods. **fig. S10.** Sampling locations of European freshwater sites included in this study.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.156215.suppl1

Supplementary material 2

The methods section of the paper

Authors: Julia K. Nuy

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.156215.suppl2

Supplementary material 3

Additional information

Authors: Julia K. Nuy

Data type: xlsx

Explanation note: **table S1.** Physicochemical features of 21 lakes. **table S2.** Statistical comparison of sequencing methods across OTU- and marker-based community profiles. **table S3.** Leave-one-out cross-validation of pairwise community concordance across sequencing methods. **table S4.** Differential OTUs and classes from differential abundance analyses. **table S5.** Results of dbRDA and BioEnv analyses. **table S6.** Cross-validation results of BioEnv analyses for presence-absence and relative abundance data at OTU- and marker-based levels. **table S7.** Number of reduced taxa in datasets. **table S8.** OTU taxonomy.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.156215.suppl3